Reference: *Biol. Bull.* **182**: 97–104. (February, 1992)

The Divergence of Species-Specific Abalone Sperm Lysins is Promoted by Positive Darwinian Selection

YOUN-HO LEE AND VICTOR D. VACQUIER

Marine Biology Research Division 0202, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093

Abstract. Recognition by sperm lysin of the egg vitelline envelope may be one component in determining the species-specificity of fertilization in abalones. The amino acid sequences of lysin proteins of seven California abalone species were deduced from the cDNA sequences. This is the first extensive comparison of a gamete recognition protein from congeneric species. Each prelysin has a highly conserved signal peptide of 18 amino acids, followed by a mature sequence of 136–138 residues. Of 136 aligned positions, 68 have the same amino acid in all seven sequences. The % identity relative to the red abalone lysin sequence is: white 90%, flat 83%, pinto 82%, pink 78%, black 71%, and green 65%. Hydropathy plots and a distance tree of the seven lysins show that red, white, and flat lysins are more closely related to each other than to the lysins of the other four species. A hypervariable, species-specific, domain exists in all sequences between positions 2–12. Amino acid replacements between any two lysins are mostly nonconservative. Analysis of the cDNA sequences shows the number of nonsynonymous substitutions (amino acid altering) exceeds the number of synonymous substitutions (silent) in 20 of the 21 pairwise comparisons of the seven sequences, indicating that positive Darwinian selection must promote the divergence of lysin sequences.

Introduction

A striking feature of fertilization is the species specificity of sperm-egg interaction in mammals (O'Rand, 1988; Yanagimachi, 1988a, 1988b; Roldan and Yanagimachi, 1989) and invertebrates (Giudice, 1973; Summers and Hylander, 1975, 1976; Osanai and Kyozuka, 1982). Sperm-egg mixtures from the same species usually yield zygotes more efficiently than cross-species mixtures. Although cross-species hybrid zygotes can be obtained in mammals and invertebrates, the general observation is that much higher concentrations of sperm are needed in the insemination mixture to achieve fertilization. Blocks to cross-species fertilization can occur at four points in the process: induction of the sperm acrosome reaction by components of the egg surface, adhesion of sperm to the egg envelope, sperm penetration of the egg envelope, and fusion of sperm and egg cell membranes. In echinoderms, the greatest barrier to cross-species fertilization is the failure of sperm to adhere to the egg vitelline envelope (Summers and Hylander, 1975, 1976); in mammals it is the failure of sperm to adhere to and penetrate the egg zona pellucida (O'Rand, 1988; Yanagimachi, 1988a,b; Roldan and Yanagimachi, 1989).

The divergence of gamete recognition proteins may be important in the establishment of barriers to cross-fertilization between populations. This may be an important factor in the speciation of marine invertebrates using external fertilization. To learn how species-specific gamete recognition proteins have diverged during evolution, we have studied a protein from abalone sperm. Abalones are marine archeogastropods of the genus *Haliotis*. Approximately 70 extant species occur on coastlines of the world, eight of them on the Pacific Coast of North America. Although abalones are members of an ancient group of gastropods, the genus *Haliotis* is relatively recent, most fossils being from the Miocene (5–25 million years; Lindberg, 1991).

The abalone egg is contained within a glycoproteinaceous vitelline envelope (VE) about 0.6 μm in thickness (Lewis *et al.*, 1982). The spermatozoon possesses a relatively enormous acrosome granule (Lewis *et al.*, 1980;

Received 25 July 1991; accepted 11 October 1991.

Abbreviations: Mr, relative molecular mass; VE, vitelline envelopes of abalone eggs.

Shiroya and Sakai, 1983) containing two abundant proteins of about Mr 18,000 and 16,000 (Lewis *et al.*, 1982). During fertilization, the sperm attaches to the egg VE, the acrosome granule opens, and the two proteins are secreted. A hole in the VE, 3 μm in diameter, is created in seconds, and the sperm passes through it to fuse with the egg (Lewis *et al.*, 1982; Sakai *et al.*, 1982). Partial amino acid sequence analysis (41 residues of the NH_2-terminus) shows that the Mr 18,000 protein is not a precursor of the Mr 16,000 protein (Vacquier, unpubl.). When the purified Mr 16,000 acrosomal protein (sperm lysin) is added to eggs, the VE rapidly dissolves by a nonenzymatic mechanism; VE glycoproteins are not degraded and no new NH_2-termini are formed (Haino-Fukushima, 1974; Lewis *et al.*, 1982; Hoshi, 1985). As previously discussed (Lewis *et al.*, 1982; Hong and Vacquier, 1986; Baginski *et al.*, 1990; Vacquier *et al.*, 1990), lysin may act by competing for hydrogen and hydrophobic bonds that hold the glycoproteinaceous fibers of the VE together.

The cDNA for pink and red abalone sperm lysins had been previously cloned and sequenced (Vacquier *et al.*, 1990). To learn about the evolutionary divergence of sperm lysin in California abalones, the polymerase chain reaction was used to generate double stranded cDNA from five additional species. The analysis of the seven deduced amino acid sequences of lysin is the first extensive comparison of a gamete recognition protein in congeneric species. We were surprised to find that the divergence of the lysin sequences is promoted by positive Darwinian selection.

Materials and Methods

The seven species of abalone used in this study were: *Haliotis corrugata* (pink, M34389), *H. cracherodii* (black, M59971), *H. fulgens* (green, M59972), *H. kamtschatkana* (pinto, M59970), *H. rufescens* (red, M34388), *H. sorenseni* (white, M59968), and *H. walallensis* (flat, M59969). The GenBank cDNA sequence accession number follows the common name of each species. The testes of male abalone were removed and poly A+ RNA isolated as described (Chomczynski and Sacchi, 1987; Vacquier *et al.*, 1990). Northern blot analysis with a full length red abalone lysin cDNA as the probe, revealed a single band of hybridization of approximately 660 nucleotides in all seven species (Vacquier *et al.*, 1990; and unpubl.). Oligonucleotide primers were synthesized to the 5′ end of the previously reported red and pink cDNA sequences (primer 6; GAACAGATTACAAG*ATG*AAGCTGT; the italicized ATG being the initiation codon), and to the complementary strand of the 3′ end of the sequence adjoining the poly A tail (primer 7; TAGTAAATCTA*TTTATT*CTGGAAT, the italicized being the complement of the poly A signal sequence; Vacquier *et al.*, 1990).

Two to ten μg of poly A+ RNA were used for first strand synthesis (Frohman, 1990). The RNA was washed twice in 1 ml 80% ethanol, dried, and dissolved in 7 μl water containing 3 μl of primer 7 (30 pMol). The tube was heated to 95°C for 5 min and then placed on ice for 10 min, followed by a 5-s centrifugation (a quick spin). A reaction mixture of 10 μl was added to the tube [the 10 μl contained: 2.5 μl dNTP mix at 2 m*M* of each nucleotide; 2.0 μl 10 × RTC buffer (Frohman, 1990); 1.0 μl human placental RNase inhibitor (Promega, Madison, Wisconsin); 2.0 μl MuLV reverse transcriptase (400 units); and 2.5 μl water]. This mixture of 20 μl was incubated for 1 h at 37°C. An additional 1 μl of MuLV reverse transcriptase was then added (200 units), and the incubation continued 1 h at 45°C. Following incubation, the sample was diluted with 2 ml 0.1 × TE (1 m*M* Tris, 0.1 m*M* EDTA, pH 8.0) and concentrated to 50 μl with a Centricon-30 microconcentrator (Amicon Inc., Beverly, Massachusetts). The redilution in 2 ml 0.1 × TE and concentration to 50–100 μl was done three times. Sixteen μl of this first strand cDNA product was used for the second strand synthesis.

To each tube of 16 μl was added 33 μl of a mixture of 5 μl PCR buffer (Frohman, 1990), 2.5 μl dimethylsulfoxide, 5.0 μl dNTPs at 2 m*M* each nucleotide, 6.0 μl primer 7 in water (30 pMol), 6.0 μl primer 6 (30 pMol), and 8.5 μl water. The 49 μl volume (in a 0.5 ml tube) was heated to 95°C for 5 min and cooled slowly (1 h) to 50°C. After a quick spin, 0.5 μl of Taq polymerase was added (2.5 units, Perkin-Elmer-Cetus, Emeryville, California), the tube vortexed gently, and 50 μl mineral oil added. The tube was incubated for 15 min at 37°C followed by 40 min at 72°C. Amplification of the lysin cDNA was accomplished in a temperature cycler by 40 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 2 min. An additional 0.5 μl of the Taq enzyme was added after the first 20 cycles. Following the last cycle, the temperature was held at 72°C for 15 min. The tubes were cooled to 23°C by a quick spin in a microfuge and 1.0 μl of Klenow fragment added, and the incubation continued for 30 min at 23°C. Agarose gel electrophoresis of the 50 μl reaction mixture showed the presence of one product of amplification of approximately 650 nucleotides.

The amplified double stranded cDNA was purified either by three cycles of dilution in 2 ml of 0.1 × TE and concentration to 50–200 μl with a Centricon-30, or by separating the product by electrophoresis in 1% agarose in 0.5 × TAE (0.04 *M* Tris acetate, 0.001 *M* EDTA, pH 8.0). The 650 base pair band was excised from the gel and the cDNA purified with Prep-A-Gene (Biorad Laboratories, Richmond, California). The cDNA was quantitated by spectrophotometry and 1 μg aliquots stored in 100% ethanol at −20°C. For sequencing, 1 μg of DNA was washed twice in 1 ml 80% ethanol and dried. The DNA

```
       -18                               -1 1                                   21
RED   :  M K L L V L C I F A M M A T L A M S R - S W H Y V E P K F L N K A F E V A L K V
WHITE:   . . . . . . . . . . . . . . . . . . . - R . . . . P . . . . . . . . . . . . . .
FLAT :   . . . . . . . . . . . . . . . . . . . - R . N F . T . R E V . . . . . . . . . .
PINTO:   . . . F . . . . . . . . . . . . . . . - . . T . . Q . . . . . . . . . . . . . E
PINK :   . . . . . . . L . . . . . . . . V . . H R F R F I P H . Y I R . E . . . . . . .
BLACK:   . . . . . . . V . . . . . . . . V . . - R Y Q F . Q H Q Y I R . . . . . . . . .
GREEN:   . . . . . . W V . . . . . . V . . . . - R . T F . R Y H Y I . . . Y . . T M . I

         22                                40                                     61
RED   :  Q I I A G F D R G L V K W L R V H G R T L S T V Q K K A L Y F V N R R Y M Q T H
WHITE:   . . . . . . . . T . . . . . . T . . G . . . H . . . . . . . . . . . . . . . . .
FLAT :   . . . . . . . . T . . . . . . A . . . S . . H . . . . . . . . . . . . . . . . .
PINTO:   E . . V R W . . K . A . . . . . N . G P . . H . . . . . . . . . . . . . . . . .
PINK :   E . . . . . . . T . . . . . . . . . G R . . . . . . . . . . . . . . . . . . . .
BLACK:   E . . . . . . . T . . . . . . N . . . G . N E N . R . V . . . . . . . . . . . .
GREEN:   . . . S . . . . Q . T A . . . . . . . R . T N N . . . T . F . . . . . . . . . .

         62                                80                                    101
RED   :  W A N Y M L W I N K K I D A L G R T P V V G D Y T R L G A E I G R R I D M A Y F
WHITE:   . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . D . .
FLAT :   . . . . . . . V . . . . . . . . . . A . . A . . . . . . . . . . . . . . . D . .
PINTO:   . . . . . . . . . . . T . . . . . . . . . . . . R . . . . . . . . . . . . V F .
PINK :   . Q . . . . . . V R . T . . . . . P . . . A . . S . . . . . . . . . . . . . . .
BLACK:   . Q . . . . . . V . . T N . . . . P . . . A . . R A . . . . . . . . . . . G . .
GREEN:   . Q . . . . . V K R . . K . . . . P A A . . . . . . . . . . . . . . V . . V F .

         102                               120                            136
RED   :  Y D F L K D K N M I P K Y L P Y M E E I N R M R P A D V P V K Y M - - G K
WHITE:   . N . . . G R . . . . . . . . . . . . . . . . . S . . I . . . . . - - . .
FLAT :   . N . . . G R . . . . . . . . . . . . . . . . . S . . I . I R . R - - . .
PINTO:   . K . . S G R . . . . . . M . . . . Q . . . . . A . . I . . . . . - - . .
PINK :   . N . . N G R . . . . . . . . . . . . . . . . . . . . . . . A N R - - . .
BLACK:   . N . . . N R V . . . . . . . . . R R L . N . . . E . . . . A N R N P . .
GREEN:   . N . . S G R K . . . P . S A . . A K L . A L . . . . . . . . N H - - . .
```

Figure 1. Amino acid sequences of the seven abalone sperm lysins. Dots denote identity to the red lysin sequence, dashes are for alignment, and numbering refers to the red lysin sequence. The signal peptide spans positions −18 to −1. The single letter amino acid code is used.

pellet was dissolved in 10 μl of sequenase buffer (U. S. Biochemicals, Cleveland, Ohio) containing 10 pMol of a sequencing primer. Eight different oligonucleotide primers were used for sequencing, all of them synthesized so as to correspond to the red abalone lysin cDNA sequence (Vacquier *et al.*, 1990). The tube containing the mixture of DNA and primer was heated to 95°C for 5 min and snap frozen 5 min in a dry ice ethanol bath, then placed in an aluminum block precooled to −20°C, which was allowed to warm to 23°C over a 2-h period. Following a quick spin, the Sequenase protocol was performed and the sequences of both strands of cDNA determined a minimum of two times. The cDNA and amino acid sequences were computer aligned and listed in order of similarity using the progressive alignment and tree building program given in Feng and Doolittle (1990). Hydropathy plots with a window of seven amino acids were done by the method of Kyte and Doolittle (1982).

To determine whether amino acid replacement between any two lysin sequences conserved the class of residue, the 20 amino acids were divided into 5 classes following the structural considerations of Dickerson and Geis (1983). Synonymous and nonsynonymous nucleotide substitutions were computed by the methods of Li *et al.* (1985) and Nei and Gojobori (1986).

Results

Deduced amino acid sequences

The deduced amino acid sequences, aligned and listed in order of descending similarity (Feng and Doolittle, 1990), are presented in the single letter code in Figure 1. Assignments of the initiation methionine (M at position −18), the signal sequence of 18 amino acids, the NH_2-terminal residue of the mature lysins being arginine (R at position 1), and the COOH-terminal residue being lysine (K at position 136) have been previously presented (Vacquier *et al.*, 1990). In Figure 1, dots denote identity to the red abalone lysin sequence and dashes are for alignment. The signal sequences (positions −18 to −1) have been highly conserved during evolution and are typical of eukaryotes (von Heijne, 1985). Neither cysteines nor

sites for N-linked glycosylation are found in the mature lysins. The mature pink abalone lysin is 137 residues, black abalone lysin is 138, and the lysins of the five other species are 136 residues in length. Of 136 aligned positions, 68 (50%) have the same amino acid in all seven species. The two longest regions of perfect identity are the eight residues between positions 88–95 and the 11 residues between positions 52–62. There are four occurrences of two contiguous positively charged amino acids in all seven sequences (positions 47–48, 55–56, 71–72, and 94–95). The percent identity in amino acid residues for the 21 pairwise comparisons of the seven lysin sequences (Table I) shows the decrease in similarity progressing from red lysin to green lysin. Green abalone lysin is equally dissimilar from each of the other six lysins, the percent identity varying from 63 to 65%. The region of greatest difference among all seven sequences is the 11 residue segment comprising positions 2–12 (Fig. 1). In this region, no two species have the identical sequence. Considerable difference in charge distribution is seen in this hypervariable segment. For example, the red, flat, and pinto lysins have a net charge of +1, whereas pink abalone lysin has a net charge of +6.

Hydropathy plots and branching order

Hydropathy plots of the seven mature lysin sequences are of value in showing subtle differences throughout the sequences (Fig. 2). The plots of the hypervariable domain of positions 2–12 (shaded) are in most cases species-specific. The upper three plots (red, white, flat) are quite similar, all having a large hydrophobic domain between residues 15–30. The pinto is clearly different from the top three, this large hydrophobic domain being reduced and followed by a hydrophilic domain centered at position 30. The pink lysin has a hydrophilic domain centered at residue 60 that is more similar in shape to the one in the black and green lysins than it is to the other four species. There is a moderately hydrophobic domain at about position 70 in pink lysin shared with only the black species. The pinto and green are the only two lysins having a large hydrophobic peak close to position 100. The plots for the black and green are clearly distinct from the other species in having two large hydrophilic domains centered about positions 45 and 60. However, the plot of the black lysin shows a large hydrophilic domain around position 125 which is not present in the green. The regions of 11 (positions 52–62) and eight (positions 88–95) amino acids that are invariant in all seven sequences are shown as shaded zones (Fig. 2). Both these regions are amphipathic, being hydrophobic in the NH_2-terminal direction and hydrophilic in the COOH-terminal direction.

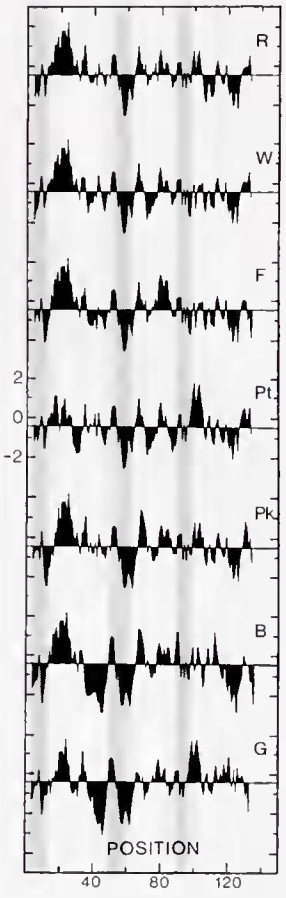

Figure 2. Hydropathy plots of the seven abalone lysin sequences using a window of seven residues. Hydrophobic values are positive and hydrophilic values negative. The hypervariable domain of 2-12, and the two invariant domains of 52-62 and 88-95 are shaded. R, red; W, white; F, flat; Pt, pinto; Pk, pink; B, black; and G, green abalone lysins.

A distance tree depicting the branching order of the seven lysin sequences (Fig. 3) shows red and white lysins to be the most closely related proteins. The black and green lysin sequences are the most divergent; they are far from the other five sequences and also far from each other.

Amino acid replacements and nucleotide substitutions

The 21 possible pairwise comparisons of the 7 lysin sequences were analyzed to determine the fraction of

Table I

Percent identity of amino acid residues in 21 pairwise comparisons in 136 aligned positions of seven lysins

Species	Red	White	Flat	Pinto	Pink	Black
White	90					
Flat	83	88				
Pinto	82	85	76			
Pink	78	80	77	72		
Black	71	72	72	65	78	
Green	65	64	65	63	65	65

Table II

The replacement of amino acids in mature lysins is nonconservative

Species	Red	White	Flat	Pinto	Pink	Black
White	$^{4}/_{13}$					
Flat	$^{12}/_{23}$	$^{10}/_{16}$				
Pinto	$^{5}/_{25}$	$^{3}/_{21}$	$^{10}/_{33}$			
Pink	$^{9}/_{29}$	$^{9}/_{26}$	$^{12}/_{30}$	$^{7}/_{37}$		
Black	$^{10}/_{40}$	$^{10}/_{38}$	$^{11}/_{38}$	$^{10}/_{47}$	$^{5}/_{27}$	
Green	$^{18}/_{48}$	$^{19}/_{49}$	$^{21}/_{47}$	$^{19}/_{51}$	$^{16}/_{46}$	$^{15}/_{48}$

amino acid replacements that were between residues of the same amino acid class (conservative replacement). The data are shown in Table II, where the numerator is the number of replacements between amino acids of the same class, and the denominator is the total number of replacements in each pairwise comparison. In all but two comparisons (flat × red and flat × white), conservative amino acid replacements are far below 50%. In summary, the majority of amino acid replacements between any two lysins involves switching the class of residue.

The number of synonymous (Ds) and nonsynonymous (Dn) nucleotide substitutions per site were computed for the 21 pairwise comparisons of the seven lysin cDNA sequences (Nei and Gojobori, 1986). The data (Table III) show that in all but one comparison (flat × green) Dn is greater than Ds. In 6 of the 21 comparisons, the difference is significant at the 5% level, and in two at the 0.5% level. These data show that positive Darwinian selection is promoting the divergence of lysin sequences. Also, the closely related sequences (Fig. 3) of red, white, flat, and pinto abalone lysins exhibit the positive selection phenomenon more strongly than do the more widely divergent sequences.

Discussion

Amino acid sequences

Homology among the seven mature lysins is readily apparent (Fig. 1). The sequences align perfectly in 952 of 955 amino acids. Lysins are constrained in length, varying from 136 to 138 amino acids. There is conservation of primary structure in that 68 of the 136 aligned positions have the identical amino acid in all seven sequences. With the exception of the hypervariable domain of positions 2–12, these invariant 68 positions are spread throughout the lysin molecule with a slight concentration toward the central portion of the sequence. Of the 68 invariant positions, 14 are occupied by residues that are highly conserved (Graur, 1985; single letter code, W = 3, G = 5, and Y = 6), and 24 by the group of seven amino acids that are replaced most frequently in mammalian proteins (Graur, 1985; T = 1, H = 1, Q = 2, F = 3, I = 4, M = 5,

Table III

Percent synonymous (Ds) and non-synonymous (Dn) nucleotide substitutions per site

Species	Ds	(SE)	Dn	(SE)	$\frac{Dn}{Ds}$	d = Dn − Ds	SE
Red × White	1.62	(1.32)	5.79	(1.40)	3.57	4.17*	1.92
Flat	2.61	(1.69)	10.08	(1.89)	3.86	7.47**	2.54
Pinto	2.76	(1.75)	10.71	(1.95)	3.88	7.95**	2.62
Pink	10.59	(3.53)	14.63	(2.33)	1.38	4.04	4.23
Black	11.08	(3.64)	21.76	(2.95)	1.96	10.68*	4.69
Green	21.27	(5.31)	24.93	(3.22)	1.17	3.66	6.21
White × Flat	4.25	(2.16)	6.36	(1.47)	1.50	2.11	2.61
Pinto	3.30	(1.91)	8.39	(1.70)	2.54	5.09*	2.56
Pink	9.92	(3.40)	13.69	(2.25)	1.38	3.77	4.08
Black	11.42	(3.69)	20.82	(2.87)	1.82	9.40*	4.67
Green	21.12	(5.27)	25.44	(3.26)	1.20	4.32	6.20
Flat × Pinto	5.45	(2.48)	13.59	(2.23)	2.49	8.14*	3.34
Pink	11.67	(3.71)	15.86	(2.45)	1.36	4.19	4.45
Black	13.42	(4.03)	22.70	(3.04)	1.69	9.28	5.05
Green	25.02	(5.84)	23.78	(3.13)	0.95	−1.24	6.63
Pinto × Pink	9.76	(3.39)	18.05	(2.63)	1.85	8.29	4.29
Black	13.71	(4.12)	26.06	(3.31)	1.90	12.35*	5.28
Green	22.66	(5.54)	24.49	(3.18)	1.08	1.83	6.39
Pink × Black	9.31	(3.28)	13.12	(2.19)	1.41	3.81	3.94
Green	16.03	(4.45)	24.02	(3.15)	1.50	7.99	5.45
Black × Green	18.14	(4.80)	27.80	(3.46)	1.53	9.66	5.92

* Significant at 5% level, ** at 0.5% level.

and L = 8). Because the occupancy of these 68 positions is identical in all 7 lysins, we conclude that they are crucial to lysin's role in fertilization in California abalones.

The hypervariable domain (positions 2–12) is strikingly similar to the ligand binding domain of annexin II (Becker *et al.*, 1990), a membrane and lipid binding protein. Annexin II possesses an NH_2-terminal 12 amino acid segment (Ac-STVHEILCKLSL) that binds its ligand (p11). Ligand binding induces the 12 residues to form a positively charged amphipathic α-helix that becomes buried in p11. The important structural features for binding between annexin II and p11 are the hydrophilic residue in position 1 and the hydrophobic side chains at positions 3, 6, 7, and 10. White abalone lysin has residues with hydrophobic side chains at positions 3, 6, 7, and 10, and the same positively charged residues at positions 4 and 9 (His and Lys) as has annexin II. In red, pink, and pinto lysins, 3 out of 4 residues at positions 3, 6, 7, and 10 have hydrophobic side chains. Three of the seven lysins are positively charged at position 4 and six out of seven at position 9. The binding of lysin to its unknown VE ligand may thus be similar to the binding of annexin II to p11 (Becker *et al.*, 1990).

Much has been learned about protein-protein recognition by X-ray crystallographic studies of the binding of proteases with their inhibitor proteins, and the binding of antibody to antigen (Janin and Chothia, 1990). In the protease-inhibitor complexes, 10–15 residues of the inhibitors make contact with 17–29 residues of the proteases. These numbers are consistent with the size of the lysin hypervariable domain. In antibodies, the antigen binding sites are disproportionately rich in residues with aromatic side chains. In the seven lysins, between positions 2–12, 26 residues of a total of 77 (34%) in all 7 sequences have aromatic side chains, whereas by total amino acid composition, only 17% of lysin residues are aromatic. This adds support to the concept that positions 2–12 in lysin may be involved in the binding of its VE ligand. We have not as yet quantitatively determined the ability of lysin to dissolve egg VEs in all 21 pairwise combinations of the 7 species. However, we have demonstrated species specificity in the cross combinations of red and pink abalone lysins and egg VEs proteins (Vacquier *et al.*, 1990).

Positive Darwinian selection in lysin divergence

In most cases, when two orthologous proteins are diverging, the frequency of synonymous (silent) nucleotide substitution (Ds) will be greater than that of nonsynonymous (amino acid altering) substitution (Dn). If positive Darwinian selection is promoting divergence of two proteins, the converse will be true. In positive selection there is adaptive value to alter the amino acid sequence. Positive selection has been proven at the molecular level in the following cases: the class I (Hughes and Nei, 1988; Hughes *et al.*, 1990) and class II (Hughes and Nei, 1989) major histocompatibility complex antigens; the V_H genes of immunoglobulins (Tanaka and Nei, 1989); the circumsporozoite antigen in *Plasmodium* (Hughes, 1991); human influenza A virus (Fitch *et al.*, 1991); and the Adh locus in *Drosophila* (McDonald and Kreitman, 1991). The nonconservative nature of amino acid replacements between lysins (Table II) provided the clue that positive selection might be promoting lysin divergence. Analysis of lysin cDNA sequences by the method of Nei and Gojobori (1986; Table III) shows that Dn exceeds Ds in 20 of 21 pairwise comparisons. Among the closely related sequences of the red, white, flat, and pinto abalone lysins (Fig. 3), Dn shows statistically significant higher values than Ds. Analysis by a similar method (Li *et al.*, 1985) yielded almost the same results. For example, in comparisons of red, white, flat, and pinto lysin sequences, the average nonsynonymous value was 2.4 times greater than the synonymous value. The most extreme comparison was between the red and flat sequences, where the nonsynonymous value was 3.4 times greater than the synonymous value. These data indicate a strong selective advantage in altering the amino acid sequence of lysin. This is the first example of positive selection acting on a gamete recognition protein. With the exception of the Adh locus in *Drosophila*, the common attribute abalone lysins share

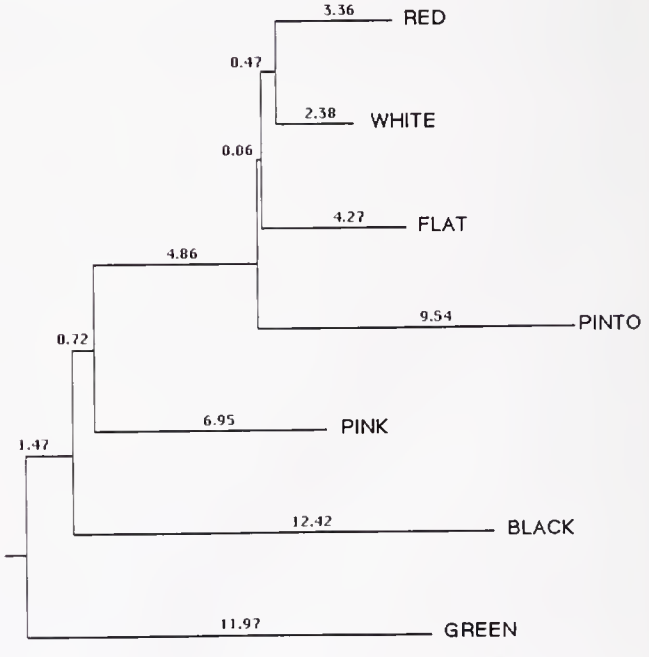

Figure 3. Distance tree showing the branching relationships of the lysin sequences. The root of the tree is placed arbitrarily at the midpoint. The numbers on the branches represent relative evolutionary distances (Feng and Doolittle, 1990).

with the published examples of positive selection is the involvement in extracellular recognition.

We cannot speculate about what could provide the selective pressure acting on lysin divergence. The demonstration of positive selection does not prove that it is a causative factor responsible for speciation in abalones. Experimental evidence exists that abalone embryos tend to settle near their parents (Prince *et al.,* 1987), and that genetic structure can occur within an abalone species in two populations separated by three km (Brown, 1991). Thus, speciation by geographic isolation probably occurs in abalones. Although the demonstration of positive selection in lysin divergence does not indicate how abalone populations split into distinct species, the possibility that it may accompany the speciation process should be considered. The statistically significant data showing positive selection (Table III) are between the closely related abalone species (Fig. 3). This suggests that a high frequency of nonsynonymous substitution (Dn) accompanies initial divergence, but Dn decreases as divergence increases. A similar situation occurs with the class I major histocompatibility genes in which Dn is greater in intralocus as compared to interlocus comparisons (Hughes and Nei, 1988). Thus, the reason that the positive selection data set is so robust for lysin may be due to the relatively recent appearance of these closely related species in the fossil record (20 million years ago; Lindberg, 1991).

One might speculate that positive selection may cause allelic variation in abalone sperm lysins. However, two male pink abalones from San Diego, California and six male red abalones (two from San Nicholas Island, two from Mendocino, California, and two from San Diego) yielded identical species-specific cDNA sequences in both the 462 nucleotide open reading frame and in about 150 nucleotides of the 3′ untranslated region containing the poly A signal sequence (Vacquier *et al.,* 1990). San Diego and Mendocino are separated by roughly 800 km of coastline, and by Point Conception, an important ecological barrier to larval transport. We tentatively conclude from these limited numbers of individuals that there is no major allelic variation in lysin sequences in the red abalone. The species-specific lysin sequences may thus be well fixed in the extant California species. As pointed out by a reviewer, there are currently no models to explain these data; they represent a genuine mystery for future research to solve.

In the class I major histocompatibility antigens, the antigen recognition site exhibits positive selection, but the different alleles share many conserved structural features making their homology obvious over tens of millions of years of evolution (Hughes and Nei, 1988). Abalone sperm lysins are similar in that strong homology exists among all seven lysins, yet 50% of the positions have species-specific amino acid replacements. Knowing the sequences of these seven sperm lysins begs the question as to the nature of sequence variation in the VE ligands that are the lysin "receptors" of the egg surface. We predict that these ligands will show the same pattern of variation; that is, some regions will be conserved in all species, while others will be hypervariable and species-specific.

Acknowledgments

We thank J. W. Swanson, R. McConnaughey, Dr. D. L. Leighton, and Dr. J. R. Pawlik for abalones, Dr. Wen-Hsiung Li for his computer program, Drs. Masatoshi Nei and Tatsuya Ota for their interest and considerable effort in preparing the data presented as Table III. Discussions with Drs. D. W. Smith, D.-F. Feng, R. F. Doolittle, W. M. Fitch, J. E. Minor, R. J. Britten, S. R. Palumbi, D. A. Powers, and F. Azam are gratefully acknowledged. Supported by NIH Grant HD12986 to V.D.V. and by a Korean Government Overseas Scholarship to Y.-H.L. This paper is dedicated to the memory of Professor Alberto Monroy.

Literature Cited

Baginsky, M. L., C. D. Stout, and V. D. Vacquier. 1990. Diffraction quality crystals of lysin from spermatozoa of the red abalone (*Haliotis rufescens*). *J. Biol. Chem.* **265:** 4958–4961.

Becker, T., K. Weber, and N. Johnson. 1990. Protein-protein recognition via short amphiphilic helices; a mutational analysis of the binding site of annexin II for p11. *EMBO J.* **9:** 4207–4213.

Brown, L. D. 1991. Genetic variation and population structure in the blacklip abalone, *Haliotis rubra*. *Aust. J. Mar. Freshwater Res.* **42:** 77–90.

Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Dickerson, R. E., and I. Geis. 1983. *Hemoglobin: structure, function, evolution and pathology.* Benjamin/Cummings Inc. Menlo Park, CA.

Feng, D.-F., and R. F. Doolittle. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. *Meth. Enzymol.* **183:** 375–387.

Fitch, W. M., J. M. E. Leiter, X. Li, and P. Palese. 1991. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* **88:** 4270–4274.

Frohman, M. A. 1990. RACE: rapid amplification of cDNA ends. Pp. 28–38 in *PCR Protocols,* M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White, eds. Academic Press, San Diego.

Giudice, G. 1973. Pp. 162–174 in *Developmental Biology of the Sea Urchin Embryo.* Academic Press, San Diego.

Graur, D. 1985. Amino acid composition and the evolutionary rates of protein-coding genes. *J. Mol. Evol* **22:** 53–62.

Haino-Fukushima, K. 1974. Studies on the egg membrane lysin of *Tegula pfeifferi:* the reaction mechanism of the egg membrane lysin. *Biochim. Biophys. Acta* **352:** 179–191.

Hong, K., and V. D. Vacquier 1986. Fusion of liposomes induced by a cationic protein from the acrosome granule of abalone spermatozoa. *Biochemistry* **25:** 543–550.

Hoshi, M. 1985. Sperm lysins. Pp. 431–462 in *Biology of Fertilization,* C. B. Metz and A. Monroy, eds. Academic Press, San Diego.

Hughes, A. L. 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127:** 345–353.

Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominant selection. *Nature* **335:** 167–170.

Hughes, A. L., and M. Nei. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* **86:** 958–962.

Hughes, A. L., T. Ota, and M. Nei. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen binding cleft of class I MHC molecules. *Mol. Biol. Evol.* **7:** 515–524.

Janin, J., and C. Chothia. 1990. The structure of protein-protein recognition sites. *J. Biol. Chem.* **265:** 16027–16030.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Lewis, C. A., D. L. Leighton, and V. D. Vacquier. 1980. Morphology of abalone spermatozoa before and after the acrosome reaction. *J. Ultrastruct. Res.* **72:** 39–47.

Lewis, C. A., C. F. Talbot, and V. D. Vacquier. 1982. A protein from abalone sperm dissolves the egg vitelline layer by a nonenzymatic mechanism. *Dev. Biol.* **92:** 227–240.

Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2:** 150–174.

Lindberg, D. R. 1991. Evolution, distribution and systematics of Haliotidae. In *Abalone of the World: Biology, Fisheries and Culture,* A. Shepherd, M. J. Tegner, and S. A. Guzman Del Proo, eds. Blackwells Scientific Publishers, London.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351:** 652–654.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonomous nucleotide substitutions. *Mol. Biol. Evol.* **3:** 418–426.

O'Rand, M. G. 1988. Sperm-egg recognition and barriers to interspecies fertilization. *Gamete Res.* **19:** 315–328.

Osanai, K., and K. Kyozuka. 1982. Cross fertilization between sea urchin eggs and oyster spermatozoa. *Gamete Res.* **5:** 49–60.

Prince, J. D., T. L. Sellers, W. B. Ford, and S. R. Talbot. 1987. Experimental evidence for limited dispersal of haliotid larvae (genus *Haliotis;* Mollusca: Gastropoda). *J. Exp. Mar. Biol. Ecol.* **106:** 243–263.

Roldan, E. R. S., and R. Yanagimachi. 1989. Cross-fertilization between Syrian and Chinese hamsters. *J. Exp. Zool.* **250:** 321–328.

Sakai, Y. T., Y. Shiroya, and K. Haino-Fukushima. 1982. Fine structural changes in the acrosome reaction of the Japanese abalone, *Haliotis discus. Dev. Growth Differ.* **24:** 531–542.

Shiroya, Y., and Y. T. Sakai. 1983. Fine structure of the spermatozoon in the Japanese abalone, *Haliotis discus. J. Wayo Women's Univ.* **24:** 253–267.

Summers, R. G., and B. L. Hylander. 1975. Species-specificity of acrosomal reaction and primary gamete binding in echinoids. *Exp. Cell Res.* **96:** 63–68.

Summers, R. G., and B. L. Hylander. 1976. Primary gamete binding: the species-exclusive event of echinoid fertilization. *Exp. Cell Res.* **100:** 190–194.

Tanaka, T., and M. Nei. 1989. Positive Darwinian selection observed at the variable region genes of immunoglobins. *Mol. Biol. Evol.* **6:** 447–459.

Vacquier, V. D., K. R. Carner, and C. D. Stout. 1990. Species-specific sequences of abalone lysin, the sperm protein that creates a hole in the egg envelope. *Proc. Natl. Acad. Sci. USA* **87:** 5792–5796.

von Heijne, G. 1985. Signal sequences: the limits of variation. *J. Mol. Biol.* **184:** 99–105.

Yanagimachi, R. 1988a. Mammalian fertilization. Pp. 135–185 in *The Physiology of Reproduction,* E. Knobil and J. Neill, eds. Raven Press, New York.

Yanagimachi, R. 1988b. Sperm-egg fusion. *Curr. Top. Membr. Transport* **32:** 3–43.